CompCytogen 12(4): 493–513 (2018)
doi: 10.3897/CompCytogen.v12i4.31015
http://compcytogen.pensoft.net

RESEARCH ARTICLE

COMPARATIVE Cytogenetics
A peer-reviewed open-access journal
International Journal of Plant & Animal Cytogenetics, Karyosystematics, and Molecular Systematics

Pilot satellitome analysis of the model plant, *Physcomitrella patens*, revealed a transcribed and high-copy IGS related tandem repeat

Ilya Kirov[1], Marina Gilyok[1], Andrey Knyazev[1], Igor Fesenko[1]

1 *Laboratory of functional genomics and proteomics of plants, Shemyakin and Ovchinnikov Institute of Bioorganic Chemistry, Moscow, Russian Federation*

Corresponding author: *Ilya Kirov* (kirovez@gmail.com)

Academic editor: *G. Karlov* | Received 2 November 2018 | Accepted 27 November 2018 | Published 13 December 2018

http://zoobank.org/62353DAD-0C5D-41F1-994B-9EDE7699F900

Citation: Kirov I, Gilyok M, Knyazev A, Fesenko I (2018) Pilot satellitome analysis of the model plant, *Physcomitrella patens*, revealed a transcribed and high-copy IGS related tandem repeat. Comparative Cytogenetics 12(4): 493–513. https://doi.org/10.3897/CompCytogen.v12i4.31015

Abstract

Satellite DNA (satDNA) constitutes a substantial part of eukaryotic genomes. In the last decade, it has been shown that satDNA is not an inert part of the genome and its function extends beyond the nuclear membrane. However, the number of model plant species suitable for studying the novel horizons of satDNA functionality is low. Here, we explored the satellitome of the model "basal" plant, *Physcomitrella patens* (Hedwig, 1801) Bruch & Schimper, 1849 (moss), which has a number of advantages for deep functional and evolutionary research. Using a newly developed pyTanFinder pipeline (https://github.com/Kirovez/pyTanFinder) coupled with fluorescence *in situ* hybridization (FISH), we identified five high copy number tandem repeats (TRs) occupying a long DNA array in the moss genome. The nuclear organization study revealed that two TRs had distinct locations in the moss genome, concentrating in the heterochromatin and knob-rDNA like chromatin bodies. Further genomic, epigenetic and transcriptomic analysis showed that one TR, named PpNATR76, was located in the intergenic spacer (IGS) region and transcribed into long non-coding RNAs (lncRNAs). Several specific features of PpNATR76 lncRNAs make them very similar with the recently discovered human lncRNAs, raising a number of questions for future studies. This work provides new resources for functional studies of satellitome in plants using the model organism *P. patens*, and describes a list of tandem repeats for further analysis.

Keywords

Physcomitrella patens, Bryophyta, satellite DNA, chromosomes, fluorescence *in situ* hybridization, long non-coding RNAs, rDNA

Copyright *Ilya Kirov et al.* This is an open access article distributed under the terms of the Creative Commons Attribution License (CC BY 4.0), which permits unrestricted use, distribution, and reproduction in any medium, provided the original author and source are credited.

Introduction

A substantial part of eukaryotic genomes is composed of different families of repetitive elements (REs). Some REs are ancient viruses (e.g., mobile elements), whereas others are *de novo* generated sequences without a specific structure. The latter include satellites, or tandem repeats (TRs), dispersed repeats and other repeat groups. TRs are the main components of heterochromatin, centromeres and telomeres (Henikoff et al. 2001, Plohl et al. 2008). TRs are important for genome stability and integrity and play a critical role in centromere function, meiotic chromosome segregation, gene regulation, X chromosome recognition and speciation (Dernburg et al. 1996, Ferree and Barbash 2009, Jagannathan et al. 2017, Menon et al. 2014, Talbert and Henikoff 2018). The genomic organization, chromosome distribution and sequence of TRs could differ significantly between closely related species and even between chromosomes of one organism (Almeida et al. 2012; Jagannathan et al. 2017, Jo et al. 2009, Kirov et al. 2017, Lim et al. 2004, Lower et al. 2018, May et al. 2005, Plohl et al. 2008, Robledillo et al. 2018, Ruiz-Ruano et al. 2016). Because TRs can mislead the recombination machine, they can also play a negative role and be the reason for genome rearrangements (Ma and Bennetzen 2006). Surprisingly, a recent study has demonstrated that TRs are not an inert part of a genome, some TRs, including those that have intergenic spacer (IGS), telomere and centromere origins, are expressed in a cell (Chen et al. 2008, May et al. 2005, Perea-Resa and Blower 2017, Yap et al. 2018, Zhao et al. 2018). Although the functions of the so-called satRNAs are enigmatic, there is a growing body of evidence that some of them can interact with different proteins and play nuclear architectural roles (Chujo et al. 2017, Staněk and Fox 2017, Sun et al. 2017, Yap et al. 2018).

The rapid evolution and high intra-monomer identity of TRs significantly hamper their study at the genome level. TRs are often collapsed or placed into an unassembled portion of the genome (e.g. Chr0, (Saint-Oyant et al. 2018)), which significantly reduces the amount of information available to study the organization of TRs. Long-read sequencing, optical mapping and other modern techniques can help to overcome these obstacles (Jain et al. 2018, Khost et al. 2017, Lower et al. 2018, Weissensteiner et al. 2017). High-throughput methods, including methods used to identify TRs from raw NGS data, have allowed researchers to gain a deeper insight into TR evolution and abundance (Lower et al. 2018, Novák et al. 2017). In addition, information about the TRs physical location is useful for understanding the TR evolution and function as well as for the improvement of the genome assembly (Saint-Oyant et al. 2018). Molecular cytogenetic techniques such as fluorescence *in situ* hybridization (FISH) or PRINS have been applied to study the genomic organization of TRs at a chromosome level (Cuadrado and Jouve 2010, Gosden et al. 1991, Jiang and Gill 2006, Kirov et al. 2017, Pavia et al. 2014, Rosato et al. 2016, Sone et al. 1999, Xiao et al. 2017). The unique nature of TRs allows their rapid localization on the chromosomes through non-denaturating FISH (ND-FISH, (Cuadrado and Jouve 2010, Jiang and Gill 2006, Kirov et al. 2017, Pavia et al. 2014, Xiao et al. 2017)). Although it is an important tool for studying the genome organization of TRs, the application of molecular cytoge-

netic methods is challenging and further improvement of chromosome preparation and FISH protocols are needed for some species (Kirov et al. 2016, Rosato et al. 2016).

The latest discoveries, including the specific transcription of some TRs as satRNAs and lncRNAs, which play important roles in regulatory processes, have moved satellite DNA biology from structural genomics to functional genomics. Satellite DNA annotation has been performed for a long list of plant species, but there are only a few model plants that are suitable for deep functional studies of TRs. In addition, no model basal plants are present on this list, although they could facilitate the study of the TR evolution mechanisms on a long timescale. Here, we performed a pilot satellitome analysis of the model basal plant, the moss *Physcomitrella patens* (Hedwig, 1801) Bruch & Schimper, 1849. It is a widely used model plant for molecular and developmental biology, evolution and biochemistry studies (van Gessel et al. 2017). The "basal" position of mosses in the land plant phylogeny makes this plant unique, bridging the gap between green algae and flowering plants (van Gessel et al. 2017). The chromosome level assembly of moss has been recently performed and different transcriptomic, epigenetic and proteomic datasets as well as tools are available (Amagai et al. 2018, Fesenko et al. 2017, Fesenko et al. 2016, Fesenko et al. 2015, Lang et al. 2005, Lang et al. 2018, Ortiz-Ramírez et al. 2016, Quatrano et al. 2007, Rensing et al. 2008, van Gessel et al. 2017). Using the newly developed pyTnaFinder pipeline (https://github.com/Kirovez/pyTanFinder), we identified five TRs that show prominent FISH signals on the nucleus and chromosomes (for two TRs). Nuclear organization revealed two TRs with distinct locations, in the heterochromatin and perinucleolar bodies. One TR, called PpNATR76, was located in the IGS of 45S rRNA genes. Using transcriptomic and genomic data, we found that PpNATR76 is transcribed into lncRNAs with unknown functions. Comparison of the distinct features of PpNATR76 organization and transcription and similarities with the recently discovered IGS-related lncRNAs in humans suggest that the transcription of a functionally important satellite containing lncRNAs from the IGS region is a conserved principle between plants and humans.

Material and methods

pyTanFinder development

pyTanFinder was written in python v3.6 using biopython (Cock et al. 2009) and networkx (Hagberg et al. 2008) libraries. Tandem Repeat Finder tool (Benson 1999) was run in the initial step of the pipeline followed by BLASTN (Altschul et al. 1997) similarity search between different monomers. Using similarity search data, the graphs were constructed by the networkx library (Hagberg et al. 2008) and a sequence with the maximum number of edges (hits) was selected for each graph. The most representative monomer sequence is then described according to its different features including accumulating abundance (the sum of the copy number of each monomer from graphs multiplied by the monomer length), monomer length and number of connections in the cluster using

matplotlib (Hunter and engineering 2007) library. The histograms are generated and represented as html document. pyTanFinder is licensed under the MIT License and is available from GitHub repository (https://github.com/Kirovez/pyTanFinder).

Slide preparation

For chromosome and nucleus preparation, the Gransden strain of *P. patens* was grown in Knop medium with 500 mg/L ammonium tartrate with 1.5% agar (Helicon, Moscow, Russian Federation) in a Sanyo Plant Growth Incubator MLR-352H (Panasonic, Osaka, Japan) with a 16-hour photoperiod at 24 °C and 61 μmol/m^2s. Gametophores at different stages (green – light green sporophyte colors) were used for analyses. Chromosome preparation was performed according to the "SteamDrop" protocol (Kirov et al. 2014) with modifications described earlier (Kirov et al. 2015). Briefly, young sporophytes were collected and fixed in Carnoy's solution (3:1, ethanol/acetic acid) for 3 h at room temperature and stored at –200 °C in 70% ethanol. The fixed material was washed twice in distilled water for 30 min and in 100 mM Citric buffer (pH 4.8). Then, the sporophytes were transferred into the enzyme mixture and incubated for 3 h at 37 °C. The 0.6% enzyme mixture containing Pectolyase Y-23 (Kikkoman, Tokyo, Japan), Cellulase Onozuka R-10 (Yakult Co. Ltd., Tokyo, Japan) and Cytohelicase (Sigma-Aldish Co.LLC, France), was prepared in 0.1 M citric buffer (pH4.8). Slides were prepared using a 1:1 (ethanol/acetic acid) mixture as the first drop and 100% acetic acid as the second drop. Then, slides were additionally incubated for 15–30 s in a drop of 60% acetic acid at 42 °C. One slide per cell suspension was checked by DAPI (100 μg/ml, 4' 6-diamidino-2-phenylindole) staining and mounted in Vectashield (Vector Laboratories, Burlingame, CA).

NGS sequencing of the moss genome

Isolated DNA was used in NGS sequencing. A sequencing library was prepared by the NEBNext ultra DNA Library Prep Kit for Illumina (New England Biolabs, UK). After preparation of the samples, the libraries were analyzed using Qubit (Invitrogen) and 2100 Bioanalyzer (Agilent Technologies). Amplification of the samples was performed according to the protocol (Illumina) using MiSeq. Raw Illumina fastq files were demultiplexed, quality filtered and analyzed using FastQC (Schmieder and Edwards 2011). RepeatExplorer tool (Novak et al. 2013) was run with default settings taking 500000 randomly selected single end reads (>100 bp) as input.

Fluorescence in situ hybridization (FISH) and microscopy

FISH was performed as previously described (Kirov et al. 2017) using TAMRA-labeled oligo probes synthesized by Evrogen (Table 1).

Table 1. Oligo probes on TRs used in FISH experiments.

ID	Sequence
17_50	(TAMRA)-AACCTTCTAGAAGAGAAGTTT
21_215	(TAMRA)-ACTTCCAGAGAGCATCGGCAA
602_86	(TAMRA)-AAGTGATGAACAAAATTTCTC
04_78	(TAMRA)-AACTTGCATTCTTCATTTTCA
592_108	(TAMRA)-ATTTCTTAGAAAATACGTTCT
20_76	(TAMRA)-AGTCCCGTCGCGAGTCCCGGA
19_95	(TAMRA)-ATAATTCTATCGGTTATGTTT
05_92	(TAMRA)-AATAATAGTAAAAGTTATAGC
21_43	(TAMRA)-ACCTTCAAGTGGACCTTAGTA
01_31	(TAMRA)-AATCAGCTCGAGTCGAGCTGA
08_44	(TAMRA)-AGCTGATGGCAGGTAAGGGAG
02_27	(TAMRA)-CTTCCGTCTTGGATCCGGAAT
08_217	(TAMRA)-AAAGTAGATCTAAAAATAAAA
05_178	(TAMRA)-ACACGAAACTCACAACTTACT
21_43	(TAMRA)-ACCTTAGTGGAGAAGTTCTGC
18_62	(TAMRA)-AGGGGAGTTTTCAAGTTTTTG
10_116	(TAMRA)-ATTGGAGAAGTATCATTGTAA
16_64	(TAMRA)-ATCGAAGAGCTAGCTTCAAGC
1004_43	(TAMRA)-AGAGAAGTTCTGTCCTTGCCT

Table 2. Primers used for qRT-PCR amplification of PpNATR76 transcripts.

Gene id	Forward	Reverse
Pp3c20_303V3.1	ATGGAGCGGGACAAGAGG	GAGTCCCGACCTCTGGCG
Pp3c20_283V3.1	CCCCCGCCAAAAATGGTTAC	CGGGACAAGGAAGAGGAGGA
Pp3c19_9271V3.1	ACTGGGCTCAAAGAAGGCAG	AGGAGGAAGAGGAGGAAGGC
Pp3c14_12290V3.2	CCCTAGCCTTTGGTTGCGTT	ACTCTCCCTTGCAATGGTCG
Pp3c4_8299V3.1	GTGTCGGGGTTAGGAAGTGG	TAGCTCTTGGAACTCGCTGC

qRT-PCR

Total RNA from protonemata tissue was isolated according to Cove et al. 2000. The RNA quality and quantity were evaluated by electrophoresis in an agarose gel with ethidium bromide staining. The exact concentration was measured using the Quant-iT RNA Assay Kit, 5–100 ng on a Qubit 3.0 (Invitrogen, US). The cDNA for RT-PCR was synthesized using the MMLV RT Kit (Evrogen, Russia). Primers (Table 2) were designed by the Primer 3.0. qRT-PCR with actin gene primer pairs was used as a positive control, whereas qRT-PCR with MQ and DNAse-treated RNA was used as a negative control. RT-PCR was performed using the qPCRmix-HS SYBR system and SYBR Green I (Evrogen) dye on a LightCycler® 96 (Roche, Mannheim, Germany). qPCR was performed in three biological and three technical replicates.

Results

Search for tandem repeats in *P. patens* genome by read clustering and pyTanFinder

To find the TRs in the *P. patens* genome, we used the Tandem Repeat Finder tool (TRF, (Benson 1999)). However, TRF provides all the TRs found in the genome; information about the copy number of individual TR monomers is unavailable. Moreover, the TRF output is redundant and it is difficult to manually handle it to find high-copy TRs that possess a certain monomer length and copy number. To overcome these obstacles we designed a python pipeline that we called pyTanFinder (https://github.com/Kirovez/pyTanFinder). It is a user-friendly command line tool to run TRF and parse the results followed by clustering of similar tandem repeats. The output of this program is a FASTA file of all tandem repeats and a table containing unique TR sequences with the estimated abundance in the genome. In addition, pyTanFinder also generates a html report containing histograms of the distribution of the TR monomer size and number of connections of each monomer into an individual cluster. We applied the pyTanFinder pipeline to the *P. patens* (v3.3) genome sequence. We identified 1518 TRs with a minimum length of genome occupy 1000 bp. Because TRs can be collapsed during genome sequence assembly, we performed low-coverage Illumina DNA sequencing followed by *de novo* annotation of TRs in next generation sequencing data using the RepeatExplorer tool (Novak et al. 2013). The clustering of the genomic reads did not reveal any clusters with a ring or globular shape that both corresponded to high-copy TRs. We then compared DNA sequences produced by the pyTanFinder pipeline and RepeatExplorer to find TRs with high copy number in both datasets. 19 TRs that were found in both datasets were used for further analysis (Table 3).

The monomer length of the TRs ranged from 27 to 217 bp (Fig.1A) and the GC content varied from 20 to 70% (Fig. 1B).

According to the pyTanFinder results, 7 (37%) TRs have high (>18000 bp, hcTRs) and 12 (63%) TRs have low (<15000 bp, lcTRs) total abundance. We were able to design primers for 5 hcTRs and obtained ladder-like or smear PCR products (Fig. 1C) that are known characteristic features of TRs (Kirov et al. 2017). Only 8 of 19 identified TRs (trTRs) were similar to the RepeatExplorer contigs from the top 200 clusters, whereas the other TRs were similar to low abundant repeat clusters. Interestingly, the pyTanFinder total abundance data did not correlate with the RepeatExplorer genome proportion data, as only 2 of the trTRs were in set of hcTRs (Table 1). Therefore, based on two approaches (pyTanFinder and RepeatExplorer) we were able to identify two sets of TRs in the moss genome that have a high and low copy number.

FISH localization of tandem repeats in *P. patens*

We used FISH to determine whether the identified TRs occupy large clusters in the moss genome. A molecular cytogenetic approach to visualize DNA sequence loci on

Table 3. General information about identified tandem repeats used for FISH analysis.

Id	Monomer length, bp	Repeat Explorer cluster	Abundancy, bp	Sequence
Pp17_50	50	10	285023	GAACCTTCTAGAAGAGAAGTTTCTAGAACCTTC TAGAAAAGAAGCCTCTG
Pp21_215	215	309	156974	CACTTCCAGAGAGCATCGGCAATTTGAACTCTC TTGTGGAGTTGAATTTGTATAGATGTCGATCCT TGAAGGCACTTCCAGAGAGCATCGGCAATTTGA ACTCTCTTGTGGAGTTGAATTTGTATGGATGTC GATCCTTGAAGGCACTTCCAGAGAGCATCGGC AATTTGAACTCTCTTGTGAAGTTGAATTTGGTA GATGTCGATCCTTGAAGG
Pp602_86	86	2626	60915	AAGTGATGAACAAAATTTCTCATTTTGCCAAGT GATGAACAAAATTTCTCATTTGCCAAGTGATGAA CAAAATTTCTCATTTTGCC
Pp04_78	78	340	38748	CAACTTGCATTCTTCATTTTCATGCTCAACTTA CATTCTCTATTTCCATGCTCAACTTGCATTCTCT ATTTCCATGCT
Pp592_108	108	1758	34258	ATTTCTTAGAAAATACGTTCTAAATGCAAAGATA CAATTTCTTAGAAAATACGTTCTAAATGCAAAG ATACAATTTCTTAGAAAATACGTTCTAAATGCAA AGATACA
Pp20_76	76	226	22386	TCCCAGTCCCGTCGCGAGTCCCGGACTTCCTC CTCCTCTTCCTTGTCCCGCCGCGACTCCCTAG TCCCGGCGCGAG
Pp19_95	95	363	18717	ATAATTCTATCGGTTATGTTTAAGGTATTCAAGA TATTATCATATACCAATGAATGAATAATGTGCCAT TGCCCACCCAAATATTGGAGTTTACC
Pp05_92	92	209	13907	CCTCTAATAATAGTAAAAGTTATAGCAATAAATAA TAATTATCAGACTTCCAATAATAGTAAAATTTATA GCAATAAATAATAATTATCGGA
Pp21_43	43	1161	10324	CCTTGCCTTCACCTTCAAGTGGACCTTAGTAGA GAAGTTTTGT
Pp01_31	31	178	5381	AATCAGCTCGAGTCGAGCTGATTTGCTTCTC
Pp08_44	44	193	3978	AGCTGATGGCAGGTAAGGGAGATTGCATGAATC AGCTCGAGTCG
Pp02_27	27	118	3648	CTTCCGTCTTGGATCCGGAATTGGCTC
Pp08_217	217	227	3472	TTTCTTAAAGTAGATCTAAAAATAAAAGTTTTGT CAAAAAAGTAGGCTTTGCTAAGTGATGACTAGA AGTGATTTCTATGTTTGAAGATGCAAAGCTCCT CTTGTTTGTTGTTAAGAAGTATAATTTACTAAAA TAAGTTATTAAATAAACAGGAAAATCAAGACGTA AGATTCCTCACAAGATTTGGGATTTACTTCAGA AAACCAACAATTCAAG
Pp05_178	718	2110	2848	CACACGAAACTCACAACTTACTCCGCACAC AACTGATCGTCGACAACGTCGTAAAGCAAG GCAACATCAGTGACAACAACGGGGAATCCT ACAGTTTTGTGTCCACAACCTTCTCCTCAC AAGTGAGATGAGGAACCCATCCGATATCTT-TGTGAGGGAGTGATGATACCGGAGGAAT
Pp21_43	43	1161	2648	GTGGACCTTAGTGGAGAAGTTCTGCCCTTGCC TTCACCTTCAA
Pp18_62	62	13	2608	AGGGGAGTTTTCAAGTTTTTGCAAGGTTACTA GTTCGGTTTCATTGGAGGTTTTTGAAGATC
Pp10_116	116	115	1619	ATTGGAGAAGTATCATTGTAAAGCAAGACTATGG AGGTATAAAAAGGGAGGTACATTTACAAGATATA GATGCCTTTGATTTAAGTTTTATTAAAAAAAAAA AAAAAAAAAAAAAA
Pp16_64	64	116	1572	GGGGTTTTTTGGATCGAAGAGCTAGCTTCAAG CTCTTTTCAAGGTCACTAGGTTGGTTTCATTA

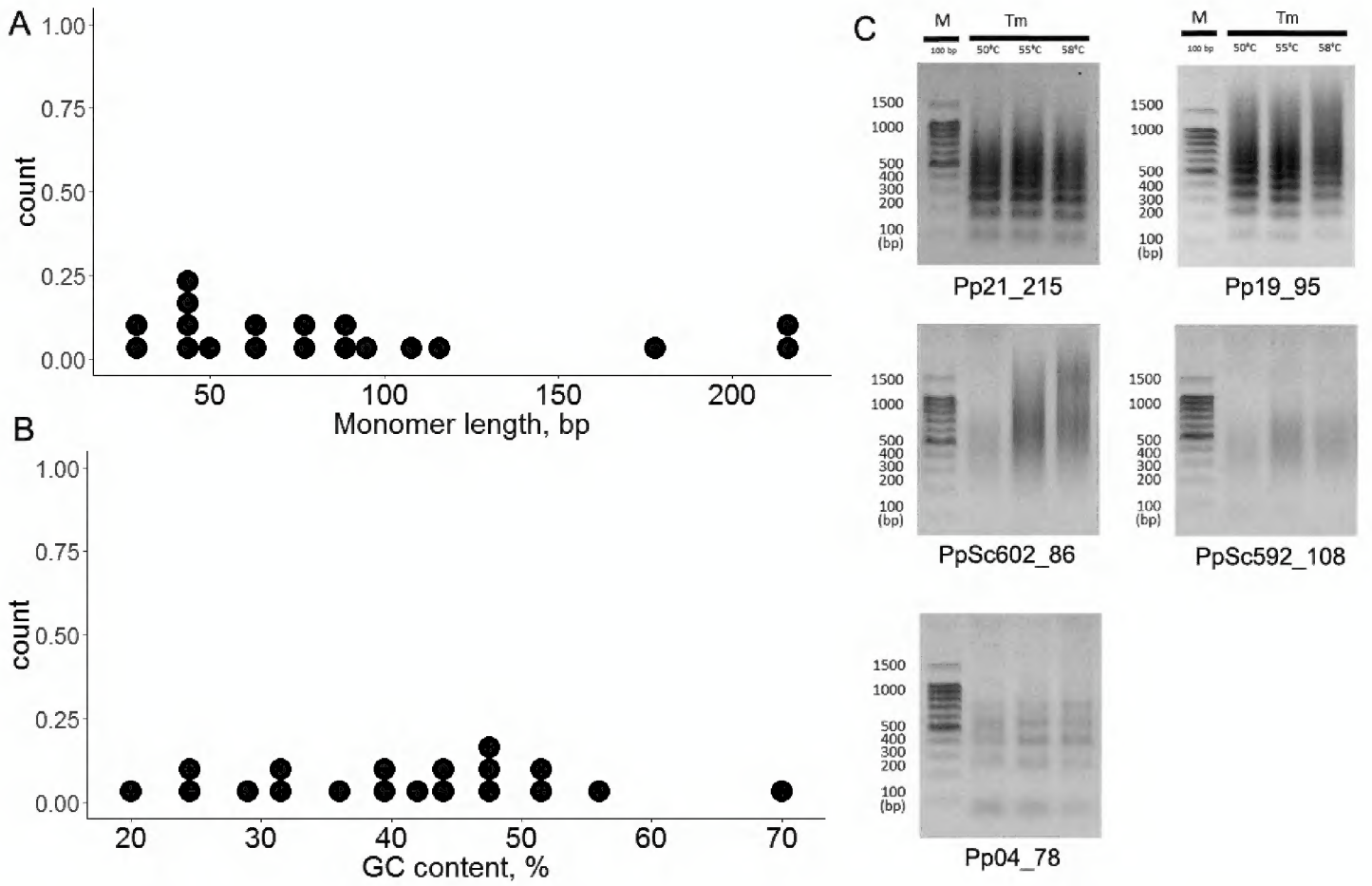

Figure 1. Features of 19 TRs. **A** Monomer length distribution **B** GC content distribution **C** Electrophoresis of PCR products from 5 TRs.

chromosomes and nuclei is challenging for bryophites (Rosato et al. 2016). To perform a pilot FISH experiment, we optimized the "SteamDrop" protocol (Kirov et al. 2014) for the preparation of the moss chromosome. Different types of material were used including protoplast, protonemata and unmatured sporophyte. No metaphase chromosomes were observed when protoplasts were used. The chromosome preparation from protonemata and unmatured sporophyte tissues resulted in a very low number of cells in the metaphase stage. Even the pretreatment of protonema tissue with different cytostatic chemicals (colchicine (3–4 h incubation in 0.05 – 0.3%), 1-bromnaphtalene (overnight incubation in saturated solution), and amiprofos-methyl (3–4 h incubation in 5 μM solution)) did not improve the results. The examples of anaphase, 1n (protonema, n=27) and 2n (sporophyte, 2n=54) metaphases as well as pachytene chromosomes after 4',6-diamidino-2-phenylindole (DAPI) staining are shown in Fig. 2.

We designed 19 TAMRA oligonucleotide probes to perform a nuclei-FISH assay. To validate that the obtained slides were suitable for FISH experiments, we used known tandemly organized sequences, Arabidopsis-type telomeric repeat ((TTTAGGG)n) and 45S rDNA, as positive controls. FISH experiments revealed many dot-like (Fig. 3A) and few distinct (Fig. 3B) signals for telomere and 45S rDNA probes, respectively, which suggested that the slides were suitable for FISH analysis in moss. We then performed nuclei-FISH experiments for 19 moss TRs. These experiments revealed 5 TRs for which FISH signals were detectable on the nuclei (Fig. 3).

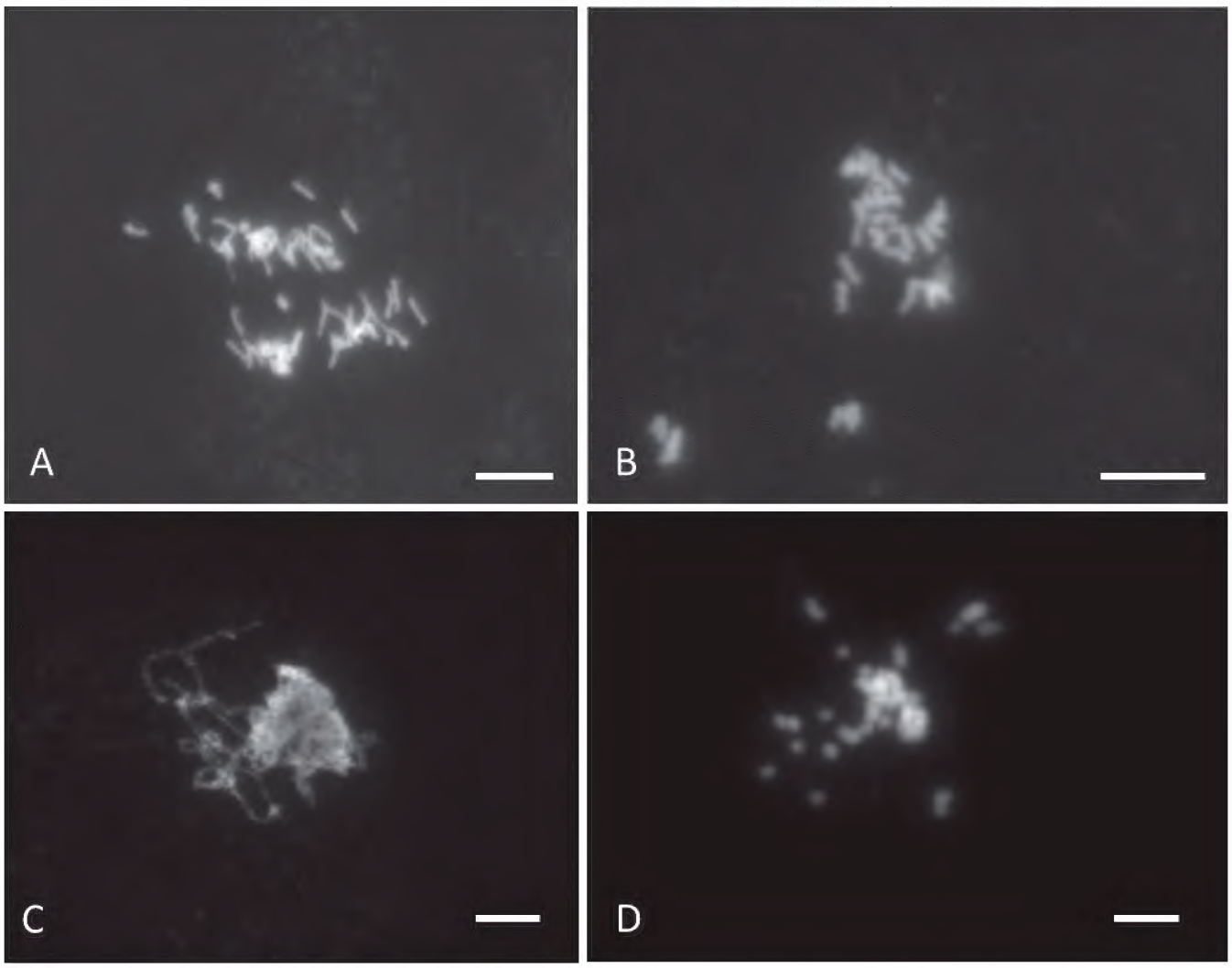

Figure 2. Mitotic and meiotic chromosomes of *P. patens* after DAPI staining. Anaphase (**A**), 1n ((**B**) protonema, n=27) and 2n ((**D**) sporophyte, 2n=54) metaphases and pachytene (**C**) stages. Scale bar: 5 μm.

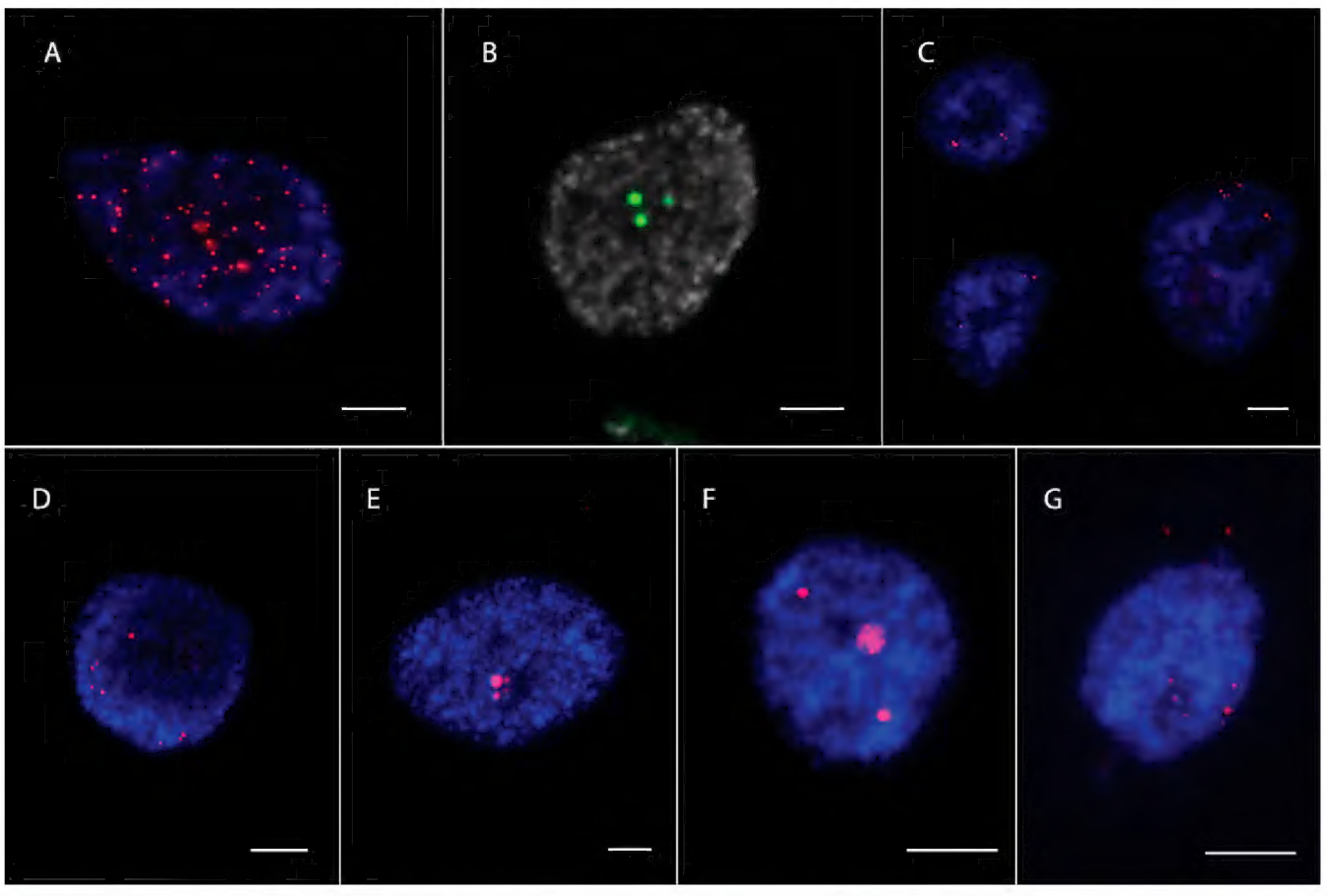

Figure 3. Results of FISH with labeled probes designed on Arabidopsis-type telomere repeat (**A**), 45S rDNA (**B**) and 5 identified TRs: Pp602_86 (**C**), Pp21_215 (**D**), Pp20_76 (**E**), Pp19_95 (**F**) and Pp592_108 (**G**).

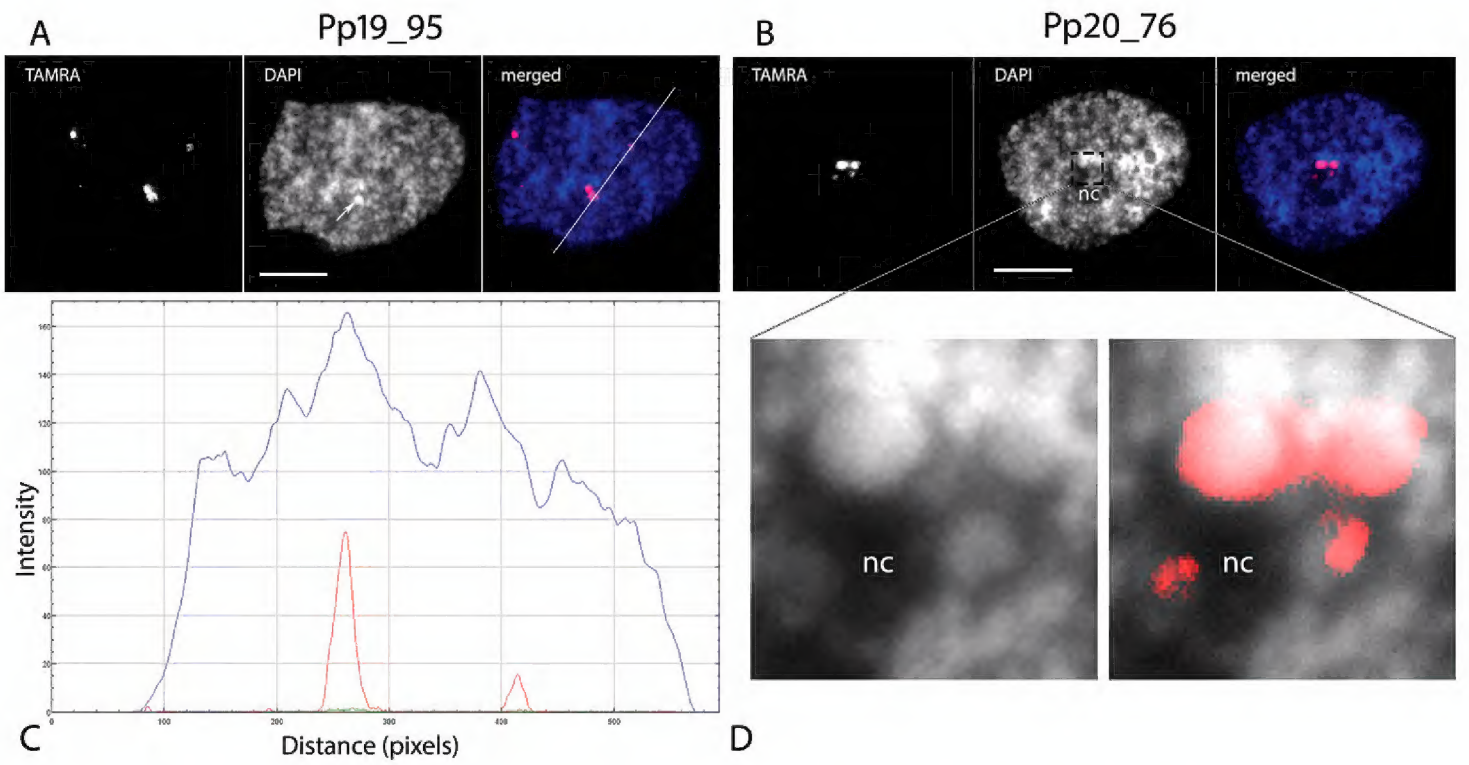

Figure 4. Nuclear organization of Pp19_95 (**A, C**) and Pp20_76 (**B, D**) TRs. **A** and **B** picture series shows fluorescence on DAPI and TAMRA channels and merged pictures **C** RGB profile of the nucleus; blue and red lines show the pixel intensity for two Pp19_95 FISH signals and DAPI staining, respectively **D** Digitally zoomed in part of the nucleus with red Pp20_76 FISH signals. nc marks the nucleolus. Scale bar: 5 μm.

Three repeats (Pp602_86, Pp21_215, Pp592_108) gave several signals that occupied two distinct territories in the nucleus. FISH signals from one TRs, Pp19_95 (95bp monomer size), were associated with heterochromatin regions of the nucleus (Fig. 4A, C) detected by DAPI. FISH signals from another TR, Pp20_76, were located at one nuclear region that was in close proximity to the nucleolus (perinucleolar region), which can be well-distinguished by DAPI staining (Fig. 4B). In contrast to Pp19_95 TR, the DAPI profile from Pp20_76 hybridization loci does not show any clear differences from neighboring nuclear regions. A closer look at the FISH signals shows that Pp20_76 loci are organized as a droplet-like structure (Fig. 4D).

Thus, nuclei FISH analysis of 19 TRs identified by pyTanFinder pipeline showed 5 TRs with pronounced signals. Moreover, one (Pp19_95) of the repeats was associated with heterochromatin structures while another one (Pp20_76) was associated with perinucleolar bodies. The 5 TRs were used for further analysis.

Location of the TRs in moss genome

To integrate our data with the *P. patens* genome sequence, we mapped 5 TRs back to the assembled *P. patens* genome sequence and estimated the genomic distribution of the TRs. Up to 45% (for Pp19_95) of BLAST hits belonged to the sequences that were not included in any assembled chromosomes (scaffolds), suggesting a challenge in the assembly of the genomic regions carrying the TRs (Fig. 5A). All BLAST hits were

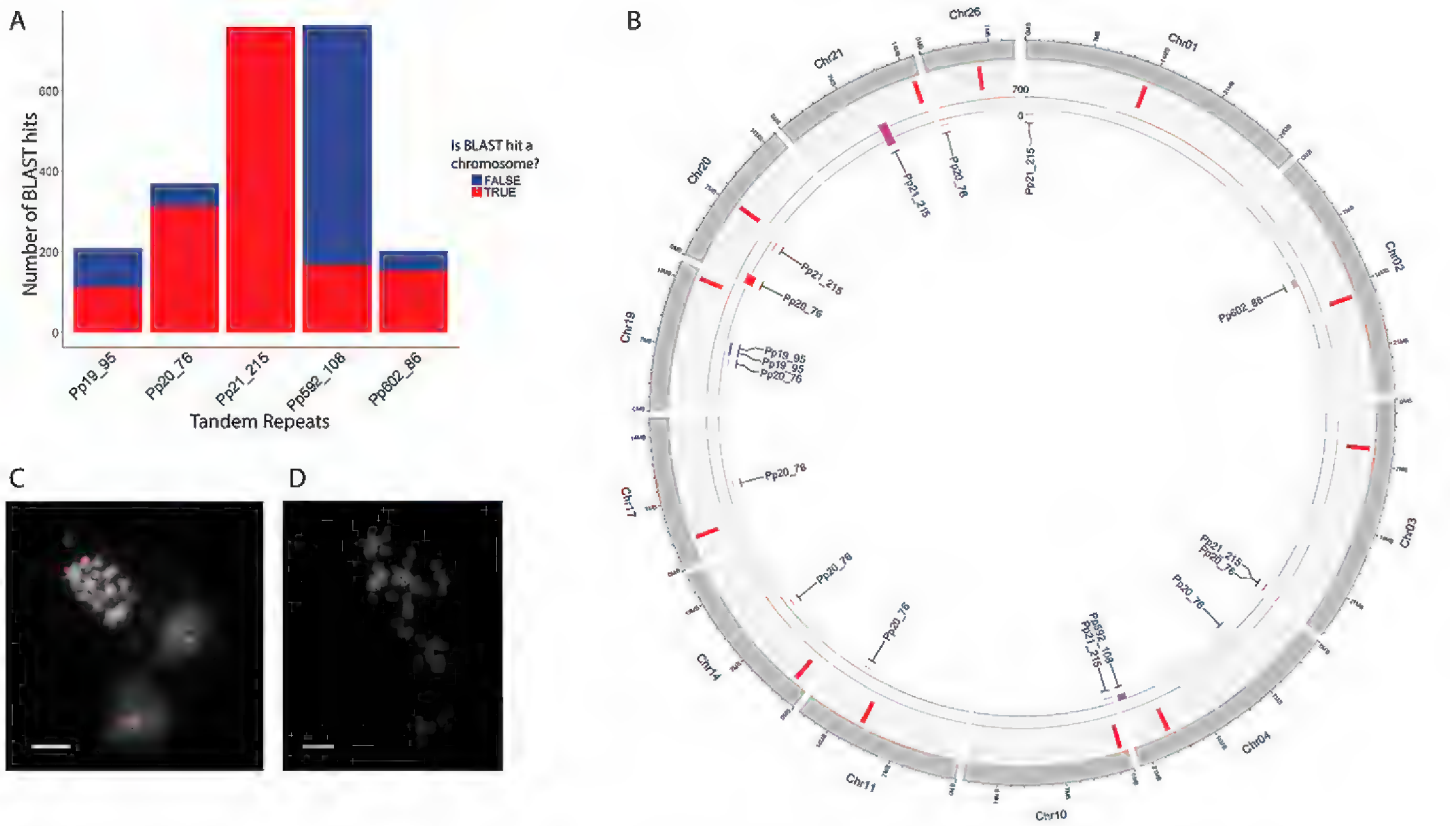

Figure 5. Chromosome location of 5 TRs. **A** Bar plot showing the number of BLAST hits derived from scaffolds and chromosome sequences **B** Circos plot: the inner layer corresponds to the bar plot showing the number of BLAST hits of the TRs on the chromosomes; FISH localization of Pp20_76 (**C**) and Pp602_86 (**D**). Scale bar: 5 μm.

distributed along 12 *P. patens* chromosomes. The Pp602_86, Pp21_215, Pp20_76, Pp19_95 and Pp592_108 TRs had 1, 5, 8, 2 and 1 loci in the assembled genome, respectively. Most of the identified loci contained only a few monomers; each of the repeats possessed a single locus with a high (up to 700) number of tandemly organized repeats including Pp21_215 (Chr21), Pp602_86 (Chr02), Pp592_108 (Chr01), Pp19_95 (Chr19) and Pp20_76 (Chr20). Two TRs, Pp21_215 and Pp20_76, had a bias toward distal parts of the chromosomes, with 60% (3) and 34% (3) loci located at the ends of the assembled chromosomes, respectively (Fig. 5B). A comparison of the putative centromere (RLC5 retrotransposon, Lang et al. 2018) and the TR locations revealed co-localization of 2 Pp21_215 (25%) loci on Chr10 and Chr20 with the RLC5-enriched regions, suggesting possible pericentromeric localization of this TR.

To further verify the results of nuclei-FISH and bioinformatics mapping, we performed FISH on moss chromosomes using two probes, Pp602_86 (single locus) and Pp20_76 (multiple loci). Although the chromosome preparation protocol needs to be further improved for *P. patens*, we were able to identify FISH signals from Pp20_76, located at the ends of two chromosome pairs, and from Pp602_86, located in the proximal positions of one chromosome pair (Fig. 5). FISH results for Pp60_86 correlated well with bioinformatics analysis which also showed a single locus on chromosome 2. In contrast, Pp20_76 has multiple loci in the moss genome assembly; two loci were revealed by FISH. One of the explanations of this discrepancy in bioinformatics and *in situ* experiments may be the limitation of FISH method sensitivity. The sensitivity of

FISH does not allow to physically map the DNA sequences if they occupy on the chromosomes less than 3–10 Kb (Valárik et al. 2004, Khrustaleva and Kik 2001). Therefore only the longest Pp20_76 array, located on Chr20, could potentially be visualized by this method. In addition, the FISH signals we observed were located at the end of the chromosomes, which is also in concordance with bioinformatics search. At the same time, a second FISH signal may be derived from Pp20_76 locus that was probably not well assembled. Therefore, the genomic mapping results together with FISH results provided evidence that the TRs that were detected occupied long clusters in the moss genome and allowed further integration of the TR location with the genomic context data available for *P. patens* (Lang et al. 2018).

Pp20_76 is located in actively transcribed chromatin

Because of the special location of Pp20_76 in the nucleus (near nucleolus) and the detected nucleus bodies enriched by this TR, we named this TR as PpNATR76(76 bp ***P. p****atens* peri**N**ucleolar **A**ssociated **T**andem **R**epeat) and analyzed it further. The alignment of 200 PpNATR76 sequences found in the moss genome showed a high conservation level between monomers. In addition, sequence analysis of the consensus PpNATR76 monomer revealed a long polypyrimidine track ($(CCT)_n$ motif). To determine why PpNATR76 DNA was located proximal to the nucleolus, we mapped the 45S rDNA to the moss genome. Using *A. thaliana* 45S rDNA gene (GenBank: X52320.1), we found two minor rDNA loci in the moss genome located on chromosomes 18 and 26 and one major rDNA locus on chromosome 20. The chromosomal location of 45S rDNA and PpNATR76 were identical on chromosomes 20 and 26, where they occupied c. 250Kb and 16Kb regions, respectively. Moreover, a detailed analysis of the loci revealed that PpNATR76 was located between 45S rDNA genes, in the IGS regions (Fig. 6A). Using the data available for moss, as a model organism, we checked the DNA and histone epigenetic landscape in the largest cluster on Chr20. We found a clear reduction in CG, CHG and CHH DNA methylation in the 45S rDNA/ PpNATR76 region (Fig. 6). In addition, the level of 'active' (H3K4me3, H3K9Ac, H3K27Ac) histone marks was significantly higher in this region compared with the flanking ones (Fig. 6). We also checked RNAseq data and found high level of RNAseq read coverage for this region, as expected for rDNA loci (Fig. 6).

PpNATR76 is transcribed into lncRNAs

Because of the transcription activity of the PpNATR7-occupying region, our next aim was to find *P. patens* transcripts possessing the PpNATR76 TR. This analysis revealed 16 transcripts whose genes were located on 5 chromosomes (Chr20, Chr19, Chr4, Chr17, Chr14). Only 4 of the transcripts possessed annotated canonical ORFs (Pp3c19_9270V3.1, p3c19_9271V3.1, Pp3c4_8299V3.1 and Pp3c14_12290V3.1). Pp3c14_12290V3.1 was the only transcript that had ORF with homology to known

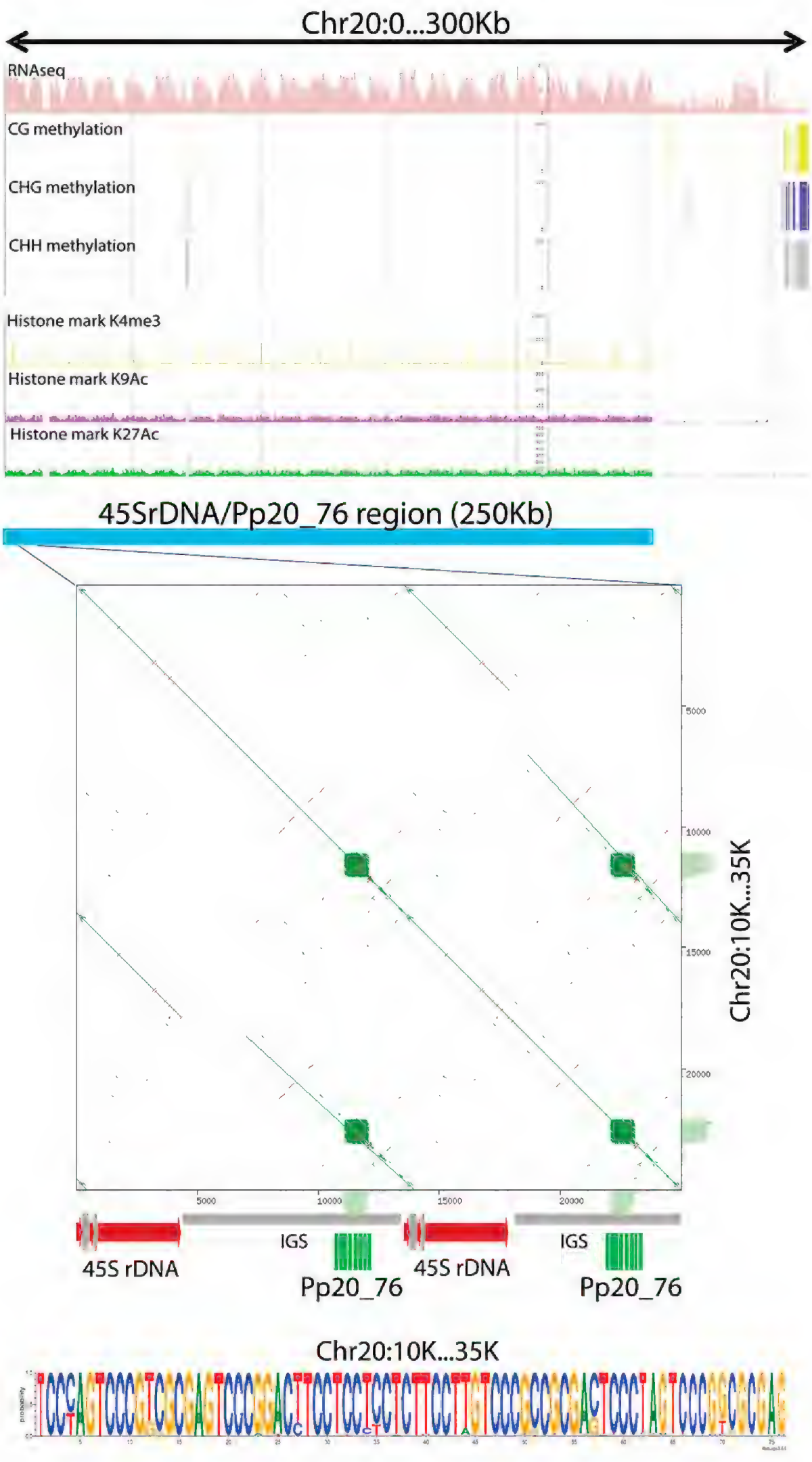

Figure 6. Genomic organization and epigenetic landscape of 45SrDNA/PpNATR76 locus. Top panel is a snapshot of CoGe GBrowser for *P. patens* (https://genomevolution.org/). Logo picture from multiple alignment of 200 PpNATR76 monomers is shown at the bottom.

proteins and was annotated as NADH:ubiquinone reductase, whereas predicted proteins from other PpNATR76 possessing transcripts did not show any homology to known proteins. These data suggested that the PpNATR76 transcripts mostly belonged to lncRNAs. To assess the robustness of the results, we performed a quantitative RT-PCR (qRT-PCR) validation of 5 PpNATR76 transcript genes (Pp3c20_303V3.1, Pp3c19_9271V3.1, Pp3c20_283V3.1, Pp3c14_12290V3.2, Pp3c4_8299V3.1) using protonemata RNA samples. For this experiment, DNA was taken as a positive control, whereas extracted RNA and MQ were negative controls. We then calculated the difference between the Cq values of pure RNA (DNA contamination control) and cDNA specific amplification. The results of qRT-PCR showed that all transcripts were expressed on detectable levels of > 5 delta. In addition, for 3 out of 5 genes, sense as well as antisense transcriptions were observed, whereas for two genes (Pp3c20_283, Pp3c14_12290) only one-way directed transcription was detected. Collectively this data proved the existence of the pPNATR76 transcripts in somatic cells and strongly suggested that PpNATR76 was transcribed as part of both protein coding and lncRNAs.

Discussion

TRs with different monomer sizes are integral parts of most eukaryotic organisms, in which they are involved in diverse biological processes. Although many efforts have been made to understand the genomic organization, structure and evolution of TRs, their functions in a cell are still poorly understood. Here, we performed a pioneering identification and FISH verification of satellite repeats, forming a long array in the genome of the model plant, *P. patens*. We developed a pipeline, pyTanFinder, and identified 19 TRs, of which 5 TRs produced FISH signals. We found both heterochromatin associated and transcribed TRs. Genomic and transcriptomic analyses identified IGS-associated moss TR, PpNATR76, which was sequestered in the perinucleolar space and transcribed as a part of lncRNAs.

pyTanFinder pipeline identified heterochromatin located satellite DNA sequences in moss

Advances in genome sequencing and bioinformatics approaches in the last decades has triggered the progress in satellite repeat isolation (reviewed by (Lower et al. 2018)). We explored the satellitome of the model plant, *P. patens*, using our pyTanFinder pipeline and repeat library generated by RepeatExplorer (Novák et al. 2013). Although a large number of TR identification tools have been developed (reviewed by (Lower et al. 2018), the pyTanFinder pipeline can be very useful if the available full genome sequence is highly fragmented. It is very common for satellite repeats to collapse during genome assembly (Saint-Oyant et al. 2018). Therefore, the identification of a TR in a single locus produced by some tools may lead to some spurious results. This limitation

is overcome in the pyTanFinder pipeline by clustering of similar TRs identified across all chromosome and scaffold sequences followed by calculation of the TR abundance based on all sequences in a cluster. This approach also makes it possible for pyTanFinder to be applied for the identification of satellite repeats in long-read single molecule real time genome sequencing data generated by modern PacBio and Oxford Nano-Pore platforms. Our preliminary results obtained on PacBio data of *Aegilops taushii* Coss., 1850 (SRA archive at NCBI: SRX3098055) supports this suggestion (data not shown). The pioneering satellite DNA identification and its FISH mapping in the moss nucleus performed in this study resulted in a set of cytogenetic markers that can be useful for future genomic and cytogenetic data integration. As shown in many other plants, the integration of chromosomal and sequence data may help to shed more light on genome evolution and to correct genome assembly ((Fransz et al. 2016, Kirov et al. 2015, Saint-Oyant et al. 2018, Shearer et al. 2014)). Molecular cytogenetic techniques, such as FISH, have never been applied to mosses; therefore, the chromosome preparation and FISH mapping procedures described in this study are important for further improvement of the *P. patens* genome assembly and annotation. Interestingly, recent (Lang et al. 2018) as well as earlier works (Melters et al. 2013) have shown low TR abundancy in the genomes of basal plants. In concordance with this observation, Lang et al. (2018) also observed a lack of clear heterochromatin regions on nuclei that typically contain TRs. Although we also did not observe large heterochromatin blocks, our slide preparation procedure allowed us to identify some small heterochromatin blocks in the moss nucleus (Figs 3, 4). In addition, the pyTanFinder pipeline allowed us to isolate at least one TR Pp19_95, which was enriched in the identified heterochromatin regions. Moreover, this repeat exhibits strong DNA methylation compared with that of the neighboring regions, which also suggested that it was located in the heterochromatin. It would be interesting to check in the future whether the heterochromatin organization is similar between basal plants and angiosperms.

Intergenic 45S rDNA spacer is a source of satellite non-coding transcripts: a principle that is conserved from first land plants to human

We found one IGS-related satellite repeat, named PpNATR76, that had several distinguishable features at the genome and transcriptome levels: 1) its DNA occupied distinct perinucleolar-associated chromatin bodies and most of its copies were located in IGS 45S rDNA spacer; 2) its DNA was hypomethylated and associated histones were enriched in 'active' chromatin marks and 3) it was transcribed into lncRNAs. The number (four signals for diploid nucleus used in this study) of PpNATR76 FISH signals was in agreement with previously observed 1–2 rDNA loci in moss and other bryophytes (Berrie 1958a, b, Rosato et al. 2016, Sone et al. 1999). As this TR was a part of the IGS region and its FISH signals on the nucleus (Fig 4B, D) were identical to 45S rDNA (Fig. 3B), we supposed that the observed PpNATR76 perinucleolar bodies were knob-like rDNA chromatin. From a first glance, this was not congruent with 'active' histone

marks and the almost absence of DNA methylation in the 45S rDNA/IGS/PpNATR76 region because the knob structure consisted of heterochromatin. However, condensed knobs and decondensed transcriptionally active rRNA genes are interspersed in one NOR region (Pontes et al. 2003). Indeed, we also found high concentration of 'inactive' chromatin marks in this region of the *P. patens* genome (H3K9me2, H3K27me3, data not shown). Because of the identity of 'active' and 'inactive' 45S rDNA sequences, the bioinformatics mapping of Chip-seq reads to the genome is not able to distinguish them and leads to erroneous results when 'active' and 'inactive' chromatin marks co-occurred. Therefore, PpNATR76 TR is a part of both knob-like ('inactive', visualized by FISH) and transcriptionally active (invisible by FISH because of the low local nuclear density of labeled loci and limited FISH sensitivity) chromatin.

Satellite DNA repeats frequently originate in plant IGS DNA and have similar organization between closely related species (Almeida et al. 2012, Falquet et al. 1997, Jo et al. 2009, Lim et al. 2004). However, the PpNATR76 length (76bp) was much shorter than the previously described IGS-associated TRs (>170 bp). IGS-associated short TRs (STR) with a monomer length range from 2 to 12 have also been described in humans (Goodwin and Swanson 2014, Yap et al. 2018). Interestingly, we showed the existence of PpNATR76 containing lncRNAs in moss cell. Recently, Yap et al. (2018) also found multiple STR-enriched lncRNAs (PNCTR) in human cell. In addition, PpNATR76 lncRNAs possess poly-pyrimidine (purine) track, which was also identified in PNCTR RNAs, where it is recognized by pyrimidine tract-binding protein (PTBP1)-specific motifs, allowing it to sequester a significant fraction of PTBP1 in the perinucleolar compartment. Poly-purine stretches were also found in another rDNA IGS-related lncRNA, PAPAS (Bierhoff H et a., 2017, Zhao et al. 2018), in which this motif is involved in forming a DNA-RNA triplex that tethers this lncRNAs to the enhancer region of rRNA genes. The described features make genomic and transcriptomic organization of moss PpNATR76 lncRNAs and human IGS related lncRNAs quite similar. Although future studies of PpNATR76 lncRNAs are required, it can be speculated that the transcription of functionally important satellite-possessing lncRNAs from the IGS region is a conserved principle between plants and humans. Because of the activity of rDNA loci, IGS-related TRs have exceptional location in the genome that promotes their transcription, resulting in the origin of novel classes of lncRNAs. This remarkable feature distinguishes this type of TR from heterochromatin-associated TRs. Our results pose a number of questions about the possible function of PpNATR76 lncRNAs as well as the existence of similar IGS-related lncRNAs in other basal species and angiosperms.

Conclusions

In this study we extended the list of model plant species for TR studies with a well-known model "basal" plant, *P. patens,* and provided a set of new FISH-verified TRs for further functional and evolutionary analysis in moss. We described a new pipeline pyTanFinder for the identification of TR in fragmented genome sequences and demon-

strated the conservation principle of IGS-related TR lncRNA expression between human and early diverged land plants. The results of our work will accelerate further studies of TR biology and function in a plant cell using the model "basal" plant *P. patens*.

Acknowledgements

This work was supported by the Russian Science Foundation (project No.17-14-01189). We thank Dr. Igor Mozhaiko for his help in moss propagation and Anna Philippova for her technical assistance in manuscript preparation.

References

Almeida C, Fonsêca A, dos Santos KGB, Mosiolek M, Pedrosa-Harand A (2012) Contrasting evolution of a satellite DNA and its ancestral IGS rDNA in Phaseolus (Fabaceae). Genome 55: 683–689. https://doi.org/10.1139/g2012-059

Altschul SF, Madden TL, Schäffer AA, Zhang J, Zhang Z, Miller W, Lipman DJ (1997) Gapped BLAST and PSI-BLAST: a new generation of protein database search programs. Nucleic Acids Research 25: 3389–3402. https://doi.org/10.1093/nar/25.17.3389

Amagai A, Honda Y, Ishikawa S, Hara Y, Kuwamura M, Shinozawa A, Sugiyama N, Ishihama Y, Takezawa D, Sakata Y (2018) Phosphoproteomic profiling reveals ABA-responsive phosphosignaling pathways in *Physcomitrella patens*. The plant Journal 94: 699–708. https://doi.org/10.1111/tpj.13891

Benson G (1999) Tandem repeats finder: a program to analyze DNA sequences. Nucleic Acids Research 27: 573–580. https://doi.org/10.1093/nar/27.2.573

Berrie GK (1958a) The Nucleolar Chromosome in Hepatics. I. Transactions of the British Bryological Society 3: 422–426. https://doi.org/10.1179/006813858804829451

Berrie GK (1958b) The Nucleolar Chromosome in Hepatics.: II. A Phylogenetic Speculation. Transactions of the British Bryological Society 3: 427–429. https://doi.org/10.1179/006813858804829334

Chen ES, Zhang K, Nicolas E, Cam HP, Zofall M, Grewal SI (2008) Cell cycle control of centromeric repeat transcription and heterochromatin assembly. Nature 451: 734. https://doi.org/10.1038/nature06561

Chujo T, Hirose T (2017) Nuclear bodies built on architectural long noncoding RNAs: unifying principles of their construction and function. Molecules and Cells 40: 889.

Cock PJ, Antao T, Chang JT, Chapman BA, Cox CJ, Dalke A, Friedberg I, Hamelryck T, Kauff F, Wilczynski B (2009) Biopython: freely available Python tools for computational molecular biology and bioinformatics. Bioinformatics 25: 1422–1423. https://doi.org/10.1093/bioinformatics/btp163

Cuadrado Á, Jouve N (2010) Chromosomal detection of simple sequence repeats (SSRs) using nondenaturing FISH (ND-FISH). Chromosoma 119: 495–503. https://doi.org/10.1007/s00412-010-0273-x

Dernburg AF, Sedat JW, Hawley RS (1996) Direct evidence of a role for heterochromatin in meiotic chromosome segregation. Cell 86: 135–146. https://doi.org/10.1016/S0092-8674(00)80084-7

Falquet J, Creusot F, Dron MJ (1997) Molecular analysis of *Phaseolus vulgaris* rDNA unit and characterization of a satellite DNA homologous to IGS subrepeats. Plant Physiology and Biochemistry 35:611–622.

Ferree PM, Barbash DA (2009) Species-specific heterochromatin prevents mitotic chromosome segregation to cause hybrid lethality in *Drosophila*. Plos Biology 7: e1000234. https://doi.org/10.1371/journal.pbio.1000234

Fesenko I, Khazigaleeva R, Kirov I, et al. (2017) Alternative splicing shapes transcriptome but not proteome diversity in *Physcomitrella patens*. Scientific Reports 7: 2698. https://doi.org/10.1038/s41598-017-02970-z

Fesenko I, Seredina A, Arapidi G, Ptushenko V, et al. (2016) The *Physcomitrella patens* chloroplast proteome changes in response to protoplastation. Frontiers in plant science 7: 1661. https://doi.org/10.3389/fpls.2016.01661

Fesenko IA, Arapidi GP, Skripnikov AY, et al. (2015) Specific pools of endogenous peptides are present in gametophore, protonema, and protoplast cells of the moss *Physcomitrella patens*. BMC Plant Biology 15: 87. https://doi.org/10.1186/s12870-015-0468-7

Fransz P, Linc G, Lee CR et al. (2016) Molecular, genetic and evolutionary analysis of a paracentric inversion in *Arabidopsis thaliana*. The Plant Journal 88: 159–178. https://doi.org/10.1111/tpj.13262

Goodwin M, Swanson MS (2014) RNA-binding protein misregulation in microsatellite expansion disorders. Systems Biology of RNA Binding Proteins. Springer: 353–388. https://doi.org/10.1007/978-1-4939-1221-6_10

Gosden J, Hanratty D, Starling J, Fantes J, Mitchell A, Porteous D (1991) Oligonucleotide-primed in situ DNA synthesis (PRINS): a method for chromosome mapping, banding, and investigation of sequence organization. Cytogenetic and Genome Research, 57: 100–104. https://doi.org/10.1159/000133122

Hagberg A, Swart PS, Chult D (2008) Exploring network structure, dynamics, and function using NetworkX. Proceedings of the 7th Python in Science Conference (SciPy2008). Pasadena, CA USA, 2008. 11–15.

Henikoff S, Ahmad K, Malik HS (2001) The centromere paradox: stable inheritance with rapidly evolving DNA. Science 293: 1098–1102. https://doi.org/10.1126/science.1062939

Hunter JD (2007) Matplotlib: A 2D graphics environment. Computing in science & engineering 9.3: 90–95. https://doi.org/10.1109/MCSE.2007.55

Jagannathan M, Warsinger-Pepe N, Watase GJ, Yamashita YM (2017) Comparative analysis of satellite DNA in the *Drosophila melanogaster* species complex. G3: Genes, Genomes, Genetics 7: 693–704. https://doi.org/10.1534/g3.116.035352

Jain M, Olsen HE, Turner DJ, Stoddart D, Bulazel KV, Paten B, Haussler D, Willard HF, Akeson M, Miga KH (2018) Linear assembly of a human centromere on the Y chromosome. Nature Biotechnology 36: 321. https://doi.org/10.1038/nbt.4109

Jiang J, Gill BS (2006) Current status and the future of fluorescence in situ hybridization (FISH) in plant genome research. Genome 49: 1057–1068. https://doi.org/10.1139/g06-076

Jo S-H, Koo D-H, Kim JF, Hur C-G, Lee S, Yang T-j, Kwon S-Y, Choi D (2009) Evolution of ribosomal DNA-derived satellite repeat in tomato genome. BMC Plant Biology 9: 42. https://doi.org/10.1186/1471-2229-9-42

Khost DE, Eickbush DG, Larracuente AM (2017) Single-molecule sequencing resolves the detailed structure of complex satellite DNA loci in *Drosophila melanogaster*. Genome Research 27: 709–721. https://doi.org/10.1101/gr.213512.116

Khrustaleva LI, Kik C (2001) Localization of single-copy T-DNA insertion in transgenic shallots (Allium cepa) by using ultra-sensitive FISH with tyramide signal amplification. The Plant Journal 25: 699–707. https://doi.org/10.1046/j.1365-313x.2001.00995.x

Kirov I, Divashuk M, Van Laere K, Soloviev A, Khrustaleva L (2014) An easy "SteamDrop" method for high quality plant chromosome preparation. Molecular Cytogenetics 7: 21. https://doi.org/10.1186/1755-8166-7-21

Kirov IV, Kiseleva AV, Van Laere K, Van Roy N, Khrustaleva LI (2017) Tandem repeats of *Allium fistulosum* associated with major chromosomal landmarks. Molecular Genetics and Genomics 292: 453–464. https://doi.org/10.1007/s00438-016-1286-9

Kirov IV, Van Laere K, Khrustaleva LI (2015) High resolution physical mapping of single gene fragments on pachytene chromosome 4 and 7 of Rosa. BMC Genetics 16: 74. https://doi.org/10.1186/s12863-015-0233-9

Kirov IV, Van Laere K, Van Roy N, Khrustaleva LI (2016) Towards a FISH-based karyotype of *Rosa* L. (Rosaceae). Comparative Cytogenetics 10: 543–554. https://doi.org/10.3897/compcytogen.v10i4.9536

Lang D, Eisinger J, Reski R, Rensing S (2005) Representation and high-quality annotation of the *Physcomitrella patens* transcriptome demonstrates a high proportion of proteins involved in metabolism in mosses. Plant Biology 7: 238–250. https://doi.org/10.1055/s-2005-837578

Lang D, Ullrich KK, Murat F et al. (2018) The *Physcomitrella patens* chromosome-scale assembly reveals moss genome structure and evolution. The Plant Journal 93: 515–533. https://doi.org/10.1111/tpj.13801

Lim K, Skalicka K, Koukalova B, Volkov R, Matyasek R, Hemleben V, Leitch A, Kovarik A (2004) Dynamic changes in the distribution of a satellite homologous to intergenic 26-18S rDNA spacer in the evolution of *Nicotiana*. Genetics 166: 1935–1946. https://doi.org/10.1534/genetics.166.4.1935

Lower SS, McGurk MP, Clark AG, Barbash DA (2018) Satellite DNA evolution: old ideas, new approaches. Current Opinion in Genetics & Development 49: 70–78. https://doi.org/10.1016/j.gde.2018.03.003

Ma J, Bennetzen JL (2006) Recombination, rearrangement, reshuffling, and divergence in a centromeric region of rice. PNAS 103: 383–388. https://doi.org/10.1073/pnas.0509810102

May BP, Lippman ZB, Fang Y, Spector DL, Martienssen RA (2005) Differential regulation of strand-specific transcripts from *Arabidopsis* centromeric satellite repeats. PLOS Genetics 1: e79. https://doi.org/10.1371/journal.pgen.0010079

Melters DP, Bradnam KR, Young HA, et al. (2013) Comparative analysis of tandem repeats from hundreds of species reveals unique insights into centromere evolution. Genome Biology 14: R10. https://doi.org/10.1186/gb-2013-14-1-r10

Menon DU, Coarfa C, Xiao W, Gunaratne PH, Meller VH (2014) siRNAs from an X-linked satellite repeat promote X-chromosome recognition in *Drosophila melanogaster*. PNAS 111: 16460–16465. https://doi.org/10.1073/pnas.1410534111

Novák P, Ávila Robledillo L, Koblížková A, Vrbová I, Neumann P, Macas J (2017) TAREAN: a computational tool for identification and characterization of satellite DNA from unassembled short reads. Nucleic Acids Research 45: e111–e111. https://doi.org/10.1093/nar/gkx257

Novák P, Neumann P, Pech J, Steinhaisl J, Macas J (2013) RepeatExplorer: a Galaxy-based web server for genome-wide characterization of eukaryotic repetitive elements from next-generation sequence reads. Bioinformatics 29: 792–793. https://doi.org/10.1093/bioinformatics/btt054

Ortiz-Ramírez C, Hernandez-Coronado M, Thamm A, Catarino B, Wang M, Dolan L, Feijó JA, Becker JD (2016) A transcriptome atlas of *Physcomitrella patens* provides insights into the evolution and development of land plants. Molecular Plant 9: 205–220. https://doi.org/10.1016/j.molp.2015.12.002

Pavia I, Carvalho A, Rocha L, Gaspar MJ, Lima-Brito J (2014) Physical location of SSR regions and cytogenetic instabilities in *Pinus sylvestris* chromosomes revealed by ND-FISH. Journal of Genetics 93: 567–571. https://doi.org/10.1007/s12041-014-0412-x

Perea-Resa C, Blower MD (2017) Satellite Transcripts Locally Promote Centromere Formation. Developmental Cell 42: 201–202. https://doi.org/10.1016/j.devcel.2017.07.017

Plohl M, Luchetti A, Meštrović N, Mantovani B (2008) Satellite DNAs between selfishness and functionality: structure, genomics and evolution of tandem repeats in centromeric (hetero) chromatin. Gene 409: 72–82. https://doi.org/10.1016/j.gene.2007.11.013

Pontes O, Lawrence RJ, Neves N, Silva M, Lee J-H, Chen ZJ, Viegas W, Pikaard CS (2003) Natural variation in nucleolar dominance reveals the relationship between nucleolus organizer chromatin topology and rRNA gene transcription in *Arabidopsis*. PNAS 100: 11418–11423. https://doi.org/10.1073/pnas.1932522100

Quatrano RS, McDaniel SF, Khandelwal A, Perroud P-F, Cove DJ (2007) *Physcomitrella patens*: mosses enter the genomic age. Current Opinion in Plant Biology 10: 182–189. https://doi.org/10.1016/j.pbi.2007.01.005

Rensing SA, Lang D, Zimmer AD, et al. (2008) The *Physcomitrella* genome reveals evolutionary insights into the conquest of land by plants. Science 319: 64–69. https://doi.org/10.1126/science.1150646

Robledillo LÁ, Koblížková A, Novák P, Böttinger K, Vrbová I, Neumann P, Schubert I, Macas J (2018) Satellite DNA in Vicia faba is characterized by remarkable diversity in its sequence composition, association with centromeres, and replication timing. Scientific Reports 8: 5838. https://doi.org/10.1038/s41598-018-24196-3

Rosato M, Kovařík A, Garilleti R, Rosselló JA (2016) Conserved organisation of 45S rDNA sites and rDNA gene copy number among major clades of early land plants. PLOS one 11: e0162544. https://doi.org/10.1371/journal.pone.0162544

Ruiz-Ruano FJ, López-León MD, Cabrero J, Camacho JPM (2016) High-throughput analysis of the satellitome illuminates satellite DNA evolution. Scientific Reports 6: 28333. https://doi.org/10.1038/srep28333

Saint-Oyant LH, Ruttink T, Hamama L, et al. (2018) A high-quality genome sequence of Rosa chinensis to elucidate ornamental traits. Nature Plants 4(7): 473–484. https://doi.org/10.1038/s41477-018-0166-1

Schmieder R, Edwards R (2011) Quality control and preprocessing of metagenomic datasets. Bioinformatics 27: 863–864. https://doi.org/10.1093/bioinformatics/btr026

Shearer LA, Anderson LK, De Jong H, Smit S, Goicoechea JL, Roe BA, Hua A, Giovannoni JJ, Stack SM (2014) Fluorescence in situ hybridization and optical mapping to correct scaffold arrangement in the tomato genome. G3: Genes, Genomes, Genetics, g3-114. https://doi.org/10.1534/g3.114.011197

Sone T, Fujisawa M, Takenaka M, Nakagawa S, Yamaoka S, Sakaida M, Nishiyama R, Yamato KT, Ohmido N, Fukui K (1999) Bryophyte 5S rDNA was inserted into 45S rDNA repeat units after the divergence from higher land plants. Plant Molecular Biology 41: 679–685. https://doi.org/10.1023/A:1006398419556

Staněk D, Fox AH (2017) Nuclear bodies: news insights into structure and function. Current Opinion in Cell Biology 46: 94–101. https://doi.org/10.1016/j.ceb.2017.05.001

Sun Q, Hao Q, Prasanth KV (2017) Nuclear long noncoding RNAs: key regulators of gene expression. Trends in Genetics 34 (2):142–157. https://doi.org/10.1016/j.tig.2017.11.005

Talbert PB, Henikoff SJTiG (2018) Transcribing Centromeres: Noncoding RNAs and Kinetochore Assembly. Trends in Genetics 34 (8): 587–599. https://doi.org/10.1016/j.tig.2018.05.001

Valárik M, Bartoš J, Kovářová P, Kubaláková M, De Jong JH, Doležel J (2004) High-resolution FISH on super-stretched flow-sorted plant chromosomes. The Plant Journal 37(6): 940–950. https://doi.org/10.1111/j.1365-313X.2003.02010.x

van Gessel N, Lang D, Reski R (2017) Genetics and Genomics of *Physcomitrella patens*. Plant Cell Biology 20: 1–32.

Weissensteiner MH, Pang AW, Bunikis I, Höijer I, Vinnere-Pettersson O, Suh A, Wolf JB (2017) Combination of short-read, long-read and optical mapping assemblies reveals large-scale tandem repeat arrays with population genetic implications. Genome Research 27, 697–708. https://doi.org/10.1101/gr.215095.116

Xiao Z, Tang S, Qiu L, Tang Z, Fu S (2017) Oligonucleotides and ND-FISH displaying different arrangements of tandem repeats and identification of *Dasypyrum villosum* chromosomes in wheat backgrounds Molecules 22: E973. https://doi.org/10.3390/molecules22060973

Yap K, Mukhina S, Zhang G, Tan JS, Ong HS, Makeyev EV (2018) A Short Tandem Repeat-Enriched RNA Assembles a Nuclear Compartment to Control Alternative Splicing and Promote Cell Survival. Molecular Cell 72(3): 525–540. https://doi.org/10.1016/j.molcel.2018.08.041

Zhao Z, Sentürk N, Song C, Grummt I (2018) lncRNA PAPAS tethered to the rDNA enhancer recruits hypophosphorylated CHD4/NuRD to repress rRNA synthesis at elevated temperatures. Genes and development 32: 836–848. https://doi.org/10.1101/gad.311688.118